

# Quantifying the risk of hemiplasy in phylogenetic inference

Rafael F. Guerrero<sup>a,b,1</sup> and Matthew W. Hahn<sup>a,b</sup>

<sup>a</sup>Department of Biology, Indiana University, Bloomington, IN 47405; and <sup>b</sup>Department of Computer Science, Indiana University, Bloomington, IN 47405

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved November 5, 2018 (received for review June 29, 2018)

**Convergent evolution—the appearance of the same character state in apparently unrelated organisms—is often inferred when a trait is incongruent with the species tree. However, trait incongruence can also arise from changes that occur on discordant gene trees, a process referred to as hemiplasy. Hemiplasy is rarely taken into account in studies of convergent evolution, despite the fact that phylogenomic studies have revealed rampant discordance. Here, we study the relative probabilities of homoplasy (including convergence and reversal) and hemiplasy for an incongruent trait. We derive expressions for the probabilities of the two events, showing that they depend on many of the same parameters. We find that hemiplasy is as likely—or more likely—than homoplasy for a wide range of conditions, even when levels of discordance are low. We also present a method to calculate the ratio of these two probabilities (the “hemiplasy risk factor”) along the branches of a phylogeny of arbitrary length. Such calculations can be applied to any tree to identify when and where incongruent traits may be due to hemiplasy.**

homoplasy | convergence | comparative methods | multispecies coalescent

**C**onvergent traits found in distantly related organisms are prime examples of the role of natural selection in evolution. They are often used as evidence for the importance of adaptation in shaping organismal form and function, although they may also reflect underlying constraints on developmental pathways (1). Understanding how often convergent evolution occurs, and the conditions under which it occurs, will allow us to better understand its causes. The identification of clear cases of convergence—or more broadly, homoplasy, which includes reversals to ancestral states (2)—also enables us to determine how often convergent phenotypes are underlain by convergent molecular changes (3, 4).

Homoplasy, whether due to convergence or reversal, is inferred when multiple independent evolutionary changes are required to explain the same character states observed among sampled lineages. Here, we consider cases of convergence where the same derived state has apparently evolved from the same ancestral state multiple times; this is sometimes referred to as “parallel” evolution (cf. ref. 5). To make strong inferences about homoplasy, we require both a phylogenetic tree describing the relationships among taxa and a model of trait evolution. If either the tree or the model is incorrect, this can lead to errors in inferences about the number of transitions that have occurred (6–10). To reduce errors in species trees, genome-scale datasets have been used to generate topologies that have strong statistical support at almost all nodes (e.g., refs. 11 and 12). Although concerns remain about appropriate models of trait evolution in some cases (13), the “resolution” of species trees with phylogenomic datasets would appear to have removed the main source of error in inferring character-state transitions accurately.

However, genome-wide data have also highlighted the ubiquity of gene-tree discordance—when individual gene trees have different topologies than the species tree—even when statistical support for the species tree is high (e.g., refs. 14–17). Discordance can be due to many factors, both technical and biological. The technical causes of gene-tree discordance can include a paucity of informative sites at individual loci, misidentification of

paralogs as orthologs, and a lack of fit between the sequence data and the substitution model used to infer the tree. The biological causes of gene-tree discordance are incomplete lineage sorting (ILS) and introgression/hybridization (18), although natural selection can sometimes result in discordance (e.g., ref. 19). In the presence of any of these biological processes, individual loci can have different histories from the species tree; discordance due to either introgression or ILS in the history of extant lineages does not go away over time, so studies of both ancient and recent divergences can be affected.

Discordance presents a problem for inferences of convergent evolution because it can produce patterns of homoplasy even when none has occurred (20). In a phenomenon dubbed “hemiplasy” (21), transitions that occur on branches of discordant trees that do not exist in the species tree will generate incongruent trait patterns (Fig. 14). Incongruent traits—binary traits that cannot be explained by a single trait transition on a bifurcating species tree—are the basis for claims of convergent evolution and homoplasy. As all discordant trees have discordant branches (22, 23), hemiplasy is expected to explain a substantial number of observed incongruent trait patterns in all cases where biological discordance exists (24–27). Note that hemiplasy cannot be confused for cases of convergent evolution arising from different ancestral states (26), so we do not consider this type of convergence here.

While complications due to hemiplasy have begun to be appreciated, the relative importance of hemiplasy and homoplasy in any particular set of relationships is rarely quantified (24, 28). Here, we present a model that describes the relative probabilities of hemiplasy and homoplasy for an incongruent trait. Given even minimal amounts of gene-tree discordance, we find that hemiplasy

## Significance

**Convergent evolution provides key evidence for the action of natural selection. The process of convergence is often inferred because the same trait appears in multiple species that are not closely related. However, different parts of the genome can reveal different relationships among species, with some genes or regions uniting lineages that appear unrelated in the species tree. If changes in traits occur in these discordant regions, a false pattern of convergence can be produced (known as “hemiplasy”). Here, we provide a way to quantify the probability that hemiplasy occurs and contrast it with the probability of convergence. We find that hemiplasy is likely to explain many apparent cases of convergent evolution, even when the fraction of discordant regions is low.**

Author contributions: R.F.G. and M.W.H. designed research, performed research, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence should be addressed. Email: rafguerr@indiana.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811268115/-DCSupplemental](https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811268115/-DCSupplemental).







$$HRF = \frac{P_{e,1} + P_{e,2}}{P_{e,1} + P_{e,2} + P_{o,1} + P_{o,2}}, \quad [3]$$

where the subscripts “1” and “2” represent the two possible incongruent trait patterns among three species. Although the probabilities of hemiplasy and homoplasy can be found for phylogenetic “knots” of more than three lineages (cf. refs. 22 and 37), we confine HRFs to the three-taxon case so that we can efficiently calculate them across larger species trees (see below). We calculate the values of  $P_e$  and  $P_o$  separately for each incongruent pattern because in nonultrametric trees these may be quite different from one another (e.g., when the branches leading to species “B” and “C” are not the same length; Fig. 1A). While this means that the HRF no longer represents the probability of hemiplasy for a specific incongruent pattern, it does helpfully summarize the overall risk of hemiplasy relative to homoplasy.

The HRF is intended to highlight individual branches of phylogenetic trees along which hemiplasy may be responsible for observed incongruence, in opposition to the assumption that all incongruence is due to homoplasy. An HRF can be calculated for each internal branch of a rooted species tree, but not for tip branches. Note that given the definitions of hemiplasy and homoplasy, the HRF associated with a branch does not indicate that a character-state transition has occurred along this branch: For instance, under hemiplasy due to ILS, the relevant mutation has occurred on an earlier lineage (e.g., the one with length  $t_3$  in Fig. 1A) but has remained polymorphic through the relevant branch (e.g., the one with length  $t_2$  in Fig. 1A). Instead, HRFs identify branches of a tree where the processes leading to hemiplasy may be occurring, and around which homoplastic transitions may be incorrectly inferred. In such cases, standard ancestral state reconstruction methods will infer homoplastic substitutions on the branches neighboring the one with the high HRF—either the branches directly “above” and “below” it in the case of reversals or one branch below and one branch sister to it in the case of convergence (SI Appendix, Fig. S1).

We have implemented a package written in R (<https://www.r-project.org/>) for calculating and visualizing HRFs on larger phylogenetic trees ([github.com/guerrero/pepo](https://github.com/guerrero/pepo)). The software walks through a phylogenetic tree starting from the tips, calculating HRFs on every trio of lineages (Methods). The input species tree must have branch lengths given in units of  $2N$  generations (sometimes referred to as “coalescent units”), and a mutation rate must be specified. Multiple methods can output species trees in coalescent units (e.g., the MP-EST program; ref. 38), or—under the assumption that all gene-tree discordance is due to ILS—these lengths can be estimated for internal branches of a tree by taking the proportion of discordant trees [i.e., 1 minus the concordance factor (CF)] and solving for  $t_2 = -\log(3/2(1 - CF))$ . Because HRFs are intended to aid researchers during exploratory studies of the evolution of many different types of traits, the exact value of the mutation rate used should not be a key concern. For a reasonable estimate of the mutation rate, the HRFs calculated will instead represent the relative risk among branches of a larger tree along which hemiplasy may be occurring. The exact value of  $P_e/P_o$  can still be calculated for a specific trait of interest on a smaller portion of the tree using Eqs. 1 and 2.

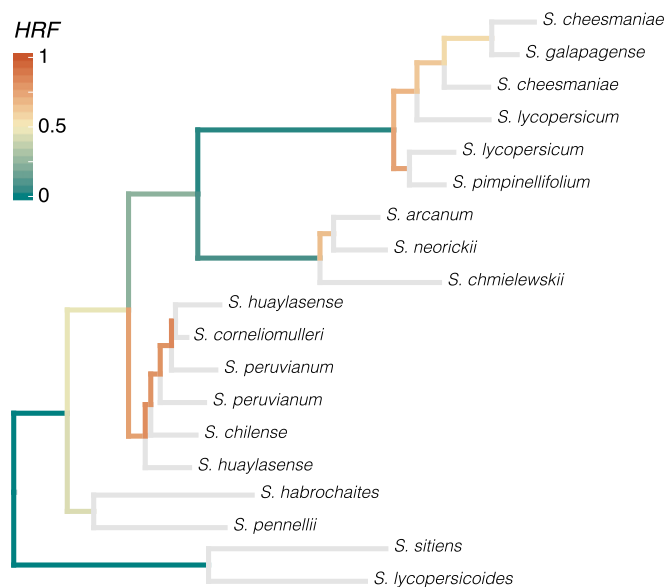
To provide an example of HRFs calculated on a larger tree, we use the phylogeny of wild tomatoes presented in Pease et al. (16). This dataset represents a rather extreme example of a recent rapid radiation, with none of the 2,745 gene trees (inferred from 100-kb windows) matching the exact topology of the species tree. After generating the necessary input species tree from a collection of gene trees (Methods), we calculated HRFs for all internal branches (Fig. 3). As can be seen, while there are multiple longer branches along which homoplasy should be a better explanation for incongruent patterns than hemiplasy, there are

also many branches with equal or higher relative probabilities of hemiplasy. The branches with high HRFs are often the shortest branches, where the most discordance is to be expected. But note that two branches with equal discordance (i.e., equal CFs; ref. 37) can have very different HRFs. The calculation of HRFs depends not only on the length of the target branch in coalescent units—which solely determines the expected degree of discordance, and to a large extent the probability of hemiplasy—but also on the length of the surrounding branches, which help to determine the probability of homoplasy. HRFs therefore represent a unique and complementary tool for understanding trait evolution on trees.

## Discussion

Phylogenomic studies have revealed high levels of gene-tree discordance in many species trees (14–17). Such discordance can be concealed by statistical measures of support for the species tree, such as bootstrap support or posterior probabilities, which can be high even when discordance is rampant. Regardless of confidence in the species tree topology, underlying gene-tree discordance means that observed patterns of trait incongruence can be due to single transitions. The phenomenon of hemiplasy likely explains the distribution of multiple ecologically important traits (e.g., refs. 39 and 40), but can mislead standard methods for inferring the number of times a trait has involved and the timing of these transitions (25–27). The goal of the work presented here is to quantify the risk of incorrectly inferring homoplasy when hemiplasy is occurring. We do this by finding explicit expressions for the probabilities of hemiplasy and homoplasy and use these expressions to develop a measure (the HRF) that highlights branches of a phylogeny along which hemiplasy may be occurring.

Our results are limited to the analytically tractable case involving three lineages and are therefore a simplification relative to larger trees. One important limitation of this approach when applying it to larger trees is that we are considering only two evolutionary events at a time: either two substitutions in the case



**Fig. 3.** Inferred tree for species in the genus *Solanum* (*S.*) labeled with HRFs. The species tree (with branch lengths in units of  $2N$  generations) was inferred by using gene trees from ref. 16; see Methods for details. HRFs were calculated for all internal branches assuming  $\mu = 0.01$  per  $2N$  generations. Values represent the proportion of incongruent traits associated with branch that are due to hemiplasy, so that HRF = 0 means no hemiplasy and HRF = 1 means all patterns due to hemiplasy.

of homoplasy or one substitution and one discordant gene tree in the case of hemiplasy. When an incongruent trait is limited to three lineages or clades (even if these clades contain multiple species), this approximation will be appropriate. Even when more than three lineages are involved, single hemiplastic substitutions can explain incongruent patterns. However, the probability of such events declines rapidly with the number of taxa involved (as the fraction of genealogies that are compatible with hemiplasy declines; *SI Appendix, Fig. S3*). Most likely, incongruent traits on rooted trees with more than five taxa will require more than two events. Such patterns may be explainable solely by homoplasy or hemiplasy, or even a mix of the two. The idiosyncrasy of larger trees and more complex patterns of incongruence likely means that simulation approaches will be necessary to address the risk of hemiplasy.

Our model is currently applicable to any type of binary trait, as these are the traits that are most often the focus of studies of convergence. While it should be straightforward to extend the model to any sort of discrete trait [e.g., gene family size (41)], even when limited to binary traits, a more important question may be what value of the mutation rate to use. We have defined the mutation rate in our model as the rate at which one state can change into the other. Although we have discussed a reversible process here (i.e., mutations from  $0 \rightarrow 1$  and  $1 \rightarrow 0$  are allowed), this is not a requirement of our model—it can accommodate other types of mutational processes as well. For instance, we could include markers for which back-mutations are thought to be impossible or rare (e.g., ref. 32) by simply removing the reversal term from Eq. 2, or by using two mutation parameters with separate rates for  $0 \rightarrow 1$  and  $1 \rightarrow 0$  transitions (equivalent to different rates in the  $\lambda_5$  and  $\lambda_2$  terms of Eq. 2, respectively). Similarly, while we have assumed that the rate of mutation is constant across the tree, separate mutation parameters for different branches can be included in future work. Fortunately, our model does not require us to specify whether the trait is molecular or phenotypic, nor any other details about the process. For studies of molecular traits—such as nucleotide substitutions—the appropriate mutation rate to use should be clear. For phenotypic traits, the mutation rate may be much higher or much lower than this per-nucleotide rate. For example, the loss of some traits may be possible by the inactivation of multiple genes (e.g., ref. 42); in such cases, the appropriate mutation rate should include the sum of nucleotide mutation probabilities at all of the sites in the genome at which inactivation can be accomplished, making it much higher than the single-site rate. In contrast, trait transitions requiring multiple nucleotide changes at a limited number of sites in the genome may have a total mutation rate that is equal to the square or cube of the per-nucleotide rate, making it even lower (and consequently making homoplasy even less likely). Although the HRF still requires a mutation rate be specified to be calculated, our hope is that it will highlight the relative risk among different branches of a larger tree regardless of the specific value used.

Our model makes two important assumptions. First, we have based all of our calculations on a coalescent model in which ILS is the only process that generates gene-tree discordance. However, at least among eukaryotic organisms, introgression appears to be a biological cause of discordance in a wide variety of systems (43). To a first approximation, the process causing discordance should not have a large effect on the results, as both ILS and introgression will be more likely to lead to discordant topologies when there are short internode branches (because introgression between sister lineages does not result in discordance). The HRF will therefore still point to the same lineages along which hemiplasy is high or low relative to homoplasy, the latter of which is not affected by introgression. In addition, our calculations assuming only ILS are likely to be underestimates of the probability of hemiplasy: If the cause of discordance is introgression, the internal branches of discordant gene trees do not have to be as short,

making the probability of mutation along them higher. On the other hand, estimates of internal branch lengths on the species tree can be artificially lowered by gene-tree discordance due to technical errors. This occurs because estimated branch lengths must accommodate observed levels of discordance, with greater discordance, implying shorter branches. As errors in gene-tree inference are more likely farther in the past, it may be that estimates of the probability of hemiplasy are commensurately less reliable for branches deeper in the species tree.

Our second important assumption is that selection is not acting on the traits of interest. Directional selection on loci underlying a trait will cause them to be more concordant relative to the neutral expectations for the case in which ILS is the sole cause of discordance. Note, however, that the magnitude of this effect may be small, especially for traits controlled by multiple loci. Even when selection on individual loci is strong, it is only selection on the internal branch of the species tree that matters—lineage-specific selection cannot affect discordance. Moreover, if introgression is the cause of hemiplasy, then selection is irrelevant to the probability of discordance. While directional selection decreases discordance, balancing selection can increase the probability of discordance. Multiple examples of hemiplasy acting on balanced polymorphisms have already been identified (e.g., refs. 14, 39, and 44). As these examples are also associated with shorter internal branches of the respective species trees, the HRF will likely also have highlighted the pertinent branches, regardless of whether the assumptions of our model have been violated.

We have modeled the probability of hemiplasy and homoplasy for a single trait. If the trait under consideration is the allelic state at a specific nucleotide—as in the primate example used above—or a phenotype controlled by such a substitution, then one may observe multiple traits showing the same incongruent pattern for at least two reasons. First, the fact that different loci across a genome will share the same discordant topology means that the appearance of multiple traits with the same pattern of incongruence are expected under hemiplasy. For this case, the assumption of independence among the single traits should be approximately correct. Second, different sites located close together in the genome will show the same incongruent pattern because of linkage. Although such stretches may be short (on the order of exon length; refs. 28 and 45), two sites in the same gene will be more likely to show the same pattern of incongruence because they share the same gene-tree topology. While such clustering of hemiplastic sites may help to identify the cause of incongruence (because homoplasy sites are not expected to be affected by linkage), they also violate our assumption of independence among traits. Further work will need to be done to fully incorporate linkage, and the correlation it induces among multiple traits, into our model.

Recent work on the genetic basis of convergent traits has revealed that such traits are sometimes determined by convergent molecular changes (e.g., refs. 46–48). The proportion of cases in which convergent phenotypes are underlain by convergent genotypes remains an open question (3, 4), and both additional phylogenetic and functional work will be needed to accurately estimate this proportion. The work here aims to aid phylogenetic studies of convergent evolution by quantifying the proportion of time the apparent convergence in traits (which would also appear to be molecular convergence) may instead be caused by underlying gene-tree discordance. Although the ecological conditions under which hemiplasy occurs may still be informative about the processes driving similarity in traits (e.g., ref. 49), properly distinguishing between hemiplasy and homoplasy is necessary for understanding the molecular basis for convergent evolution.

## Methods

**Calculating  $P_e/P_o$  for Primates.** To calculate  $P_e/P_o$  in the HC ancestor, we estimated branch lengths in coalescent units (6.8, 6.8, 7.6, 0.6, and 3.6 for H, C, G, HC, and HCG, respectively) that agree with published divergence times

and effective population sizes (34, 35, 50). These lengths assume that H and C diverged 4.1 Mya, that HC and G diverged 5.5 Mya, and that generation time is 20 y throughout the clade. Additionally, we assume effective population sizes of 15,000; 15,000; 18,000; 19,000; 65,000; and 45,000 individuals for H, C, G, HC, and HCG, respectively. Given these assumptions, the value of  $P_e/P_o$  for incongruent cases ( $H = 0, C = 1, G = 1$ ) and ( $H = 1, C = 0, G = 1$ ) is the same, but in *Results*, we focus on the latter for clarity. For the genome-wide average estimate of the population mutation rate, we use  $\mu = 1.2 \times 10^{-4}$  per 30,000 generations, which implies that  $N_e = 15,000$  individuals and a per-site mutation rate of  $4.0 \times 10^{-9}$  (one-third of the published estimated rate, since we are considering convergence between only two states); the estimates for CpG and non-CpG sites used the corresponding per-site rates (from ref. 36) and the same value of  $N_e$ . We assumed that the mutation rate was constant along the phylogeny.

**Calculating HRFs.** An HRF (Eq. 3) can be calculated for any branch with a sibling, an ancestor, and two daughter lineages (compare branch 4 in Fig. 1). In other words, all branches of a phylogeny—except the root and leaves—have an HRF value. The *pepo* package walks up a phylogeny (in R, a *phylo* list as defined by the *ape* package; ref. 51) and calculates HRF for each internal branch, returning the values in a new data frame (compatible with *treeio*;

ref. 52). The default HRF calculation in *pepo* allows reversals at the same rate as forward mutations, but these parameters can be specified by the user. This default method assumes that the ancestor of each focal branch has a length of at least  $8N$  generations [i.e.,  $t_3 = \max(4, x)$ , where  $x$  is the observed ancestral branch length]. This setting, which can also be modified by users, is intended to reduce the effect of our assumption that A, B, and C have coalesced by the end of  $t_3$ .

We calculated HRFs along the tomato phylogeny reported by Pease et al. (16), specifically the “best coalescent-based phylogeny from 100 replicates of MP-EST using 100-kb genome window trees” in the supplement of that work. Because MP-EST assigns an arbitrary length of nine to leaves, we modified terminal branch lengths in two ways. For species where multiple individuals were collected, we collapsed monophyletic samples into a single branch representing the species (the ancestral branch of the replicates). For species with a single sample, we assigned a terminal length of one, an equally arbitrary value that is probably closer to the evolutionary history of this clade.

**ACKNOWLEDGMENTS.** We thank the editor and two reviewers for constructive feedback. This work was supported by National Science Foundation Grant DBI-1564611.

- Losos JB (2011) Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.
- Wake DB, Wake MH, Specht CD (2011) Homoplasy: From detecting pattern to determining process and mechanism of evolution. *Science* 331:1032–1035.
- Rosenblum EB, Parent CE, Brandt EE (2014) The molecular basis of phenotypic convergence. *Annu Rev Ecol Syst* 45:203–226.
- Storz JF (2016) Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* 17:239–250.
- Zhang J, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequence level. *Mol Biol Evol* 14:527–536.
- Duchêne S, Lanfear R (2015) Phylogenetic uncertainty can bias the number of evolutionary transitions estimated from ancestral state reconstruction methods. *J Exp Zool B Mol Dev Evol* 324:517–524.
- Goldberg EE, Igić B (2008) On phylogenetic tests of irreversible evolution. *Evolution* 62:2727–2741.
- Huelsensbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. *Syst Biol* 52:131–158.
- Maddison WP (2006) Confounding asymmetries in evolutionary diversification and character change. *Evolution* 60:1743–1746.
- Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53:673–684.
- Dunn CW, et al. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Misof B, et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.
- Beaulieu JM, O'Meara BC, Donoghue MJ (2013) Identifying hidden rate changes in the evolution of a binary morphological character: The evolution of plant habit in campanulid angiosperms. *Syst Biol* 62:725–737.
- Fontaine MC, et al. (2015) Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347:1258524.
- Jarvis ED, et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol* 14:e1002379.
- Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425:798–804.
- Maddison WP (1997) Gene trees in species trees. *Syst Biol* 46:523–536.
- Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G (2007) C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol* 17:1241–1247.
- Hahn MW, Nakhleh L (2016) Irrational exuberance for resolved species trees. *Evolution* 70:7–17.
- Avise JC, Robinson TJ (2008) Hemiplay: A new term in the lexicon of phylogenetics. *Syst Biol* 57:503–507.
- Mendes FK, Hahn MW (2018) Why concatenation fails near the anomaly zone. *Syst Biol* 67:158–169.
- Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Math Biosci* 53:131–147.
- Copetti D, et al. (2017) Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proc Natl Acad Sci USA* 114:12003–12008.
- Mendes FK, Hahn MW (2016) Gene tree discordance causes apparent substitution rate variation. *Syst Biol* 65:711–721.
- Mendes FK, Hahn Y, Hahn MW (2016) Gene tree discordance can generate patterns of diminishing convergence over time. *Mol Biol Evol* 33:3299–3307.
- Wu M, Kostyun JL, Hahn MW, Moyle LC (2018) Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol Ecol* 27:3301–3316.
- Mendes FK, Livera A, Hahn MW (2018) The perils of intralocus recombination for inferences of molecular convergence. *Philos Trans R Soc Lond B Biol Sci*, in press.
- Hudson RR (1983) Testing the constant-rate neutral allele model with protein-sequence data. *Evolution* 37:203–217.
- Pollard DA, Iyer VN, Moses AM, Eisen MB (2006) Widespread discordance of gene trees with species tree in *Drosophila*: Evidence for incomplete lineage sorting. *PLoS Genet* 2:e173.
- Scally A, et al. (2012) Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175.
- Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol* 13:e1002224.
- Pease JB, Hahn MW (2013) More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* 67:2376–2384.
- Hobolth A, Christensen OF, Mailund T, Schierup MH (2007) Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* 3:e7.
- Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T (2011) Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res* 21:349–356.
- Kong A, et al. (2012) Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* 488:471–475.
- Ané C, Larget B, Baum DA, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Mol Biol Evol* 24:412–426.
- Liu L, Yu L, Edwards SV (2010) A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol* 10:302.
- Lamichanay S, et al. (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352:470–474.
- Li Y-Y, Liu Z, Qi F-Y, Zhou X, Shi P (2017) Functional effects of a retained ancestral polymorphism in *Prestin*. *Mol Biol Evol* 34:88–92.
- Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res* 15:1153–1160.
- Zufall RA, Rausher MD (2004) Genetic changes associated with floral adaptation restrict future evolutionary potential. *Nature* 428:847–850.
- Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *Bioessays* 38:140–149.
- Palesch D, et al. (2018) Sooty mangabey genome sequence reveals new insights into mechanisms of disease resistance in natural hosts of SIV. *Nature* 553:77–81.
- Slatkin M, Pollack JL (2006) The concordance of gene trees and species trees at two linked loci. *Genetics* 172:1979–1984.
- Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in pigmentation at multiple levels: Mutations, genes and function. *Philos Trans R Soc Lond B Biol Sci* 365:2439–2450.
- Projecto-García J, et al. (2013) Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc Natl Acad Sci USA* 110:20669–20674.
- Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P (2012) Parallel molecular evolution in an herbivore community. *Science* 337:1634–1637.
- Lee KM, Coop G (2017) Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* 207:1591–1619.
- Prado-Martinez J, et al. (2013) Great ape genetic diversity and population history. *Nature* 499:471–475.
- Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY (2017) ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36.